

The ToxBank Data Warehouse: Supporting the Replacement of In Vivo Repeated Dose Systemic Toxicity Testing

Citation for published version (APA):

Kohonen, P., Benfenati, E., Bower, D., Ceder, R., Crump, M., Cross, K., Grafström, R. C., Healy, L., Helma, C., Jeliaskova, N., Jeliaskova, V., Maggioni, S., Miller, S., Myatt, G., Rautenberg, M., Stacey, S., Willighagen, E., Wiseman, J., & Hardy, B. (2013). The ToxBank Data Warehouse: Supporting the Replacement of In Vivo Repeated Dose Systemic Toxicity Testing. *Molecular Informatics*, 32(1), 47-63. <https://doi.org/10.1002/minf.201200114>

Document status and date:

Published: 01/01/2013

DOI:

[10.1002/minf.201200114](https://doi.org/10.1002/minf.201200114)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

The ToxBank Data Warehouse: Supporting the Replacement of In Vivo Repeated Dose Systemic Toxicity Testing

Pekka Kohonen,^[a] Emilio Benfenati,^[b] David Bower,^[c] Rebecca Ceder,^[a] Michael Crump,^[c] Kevin Cross,^[c] Roland C. Grafström,^[a] Lyn Healy,^[d] Christoph Helma,^[e] Nina Jeliaskova,^[f] Vedrin Jeliaskov,^[f] Silvia Maggioni,^[b] Scott Miller,^[c] Glenn Myatt,^[c] Michael Rautenberg,^[e] Glyn Stacey,^[d] Egon Willighagen,^[a] Jeff Wiseman,^[g] and Barry Hardy^{*,[h]}

Abstract: The aim of the SEURAT-1 (Safety Evaluation Ultimately Replacing Animal Testing-1) research cluster, comprised of seven EU FP7 Health projects co-financed by Cosmetics Europe, is to generate a proof-of-concept to show how the latest technologies, systems toxicology and toxicogenomics can be combined to deliver a test replacement for repeated dose systemic toxicity testing on animals. The SEURAT-1 strategy is to adopt a mode-of-action framework to describe repeated dose toxicity, combining in vitro and in silico methods to derive predictions of in vivo toxicity responses. ToxBank is the cross-cluster infrastructure project whose activities include the development of a data warehouse to provide a web-accessible shared repository of re-

search data and protocols, a physical compounds repository, reference or "gold compounds" for use across the cluster (available via wiki.toxbank.net), and a reference resource for biomaterials. Core technologies used in the data warehouse include the ISA-Tab universal data exchange format, REpresentational State Transfer (REST) web services, the W3C Resource Description Framework (RDF) and the OpenTox standards. We describe the design of the data warehouse based on cluster requirements, the implementation based on open standards, and finally the underlying concepts and initial results of a data analysis utilizing public data related to the gold compounds.

Keywords: Toxicology • Bioinformatics • Repeated-dose toxicity testing • Alternative Testing Strategies • Toxicogenomics • Predictive toxicology • Electron transport • Redox chemistry • Oxidation • Gene expression • Data warehouse • Databases • Wikis • Cheminformatics • Cellular models • Semantic web • Reference compounds • OpenTox • SEURAT-1 • ToxBank

1 Introduction

1.1 The SEURAT-1 Cluster

The SEURAT-1 (Safety Evaluation Ultimately Replacing Animal Testing-1) research program started at the beginning of 2011 with the objective of improving safety assessment without the need for animal experiments.^[1] This €50 M project is co-financed by the European Union's FP7

Health Program^[2] and Cosmetics Europe.^[3] A critical element of this cluster is its focus on repeated dose toxicity and the adoption of a mode-of-action (MOA) framework based on an understanding of key biological events driven by levels of exposure over time.^[4–8] The SEURAT-1 cluster is comprised of five complementary research projects (COSMOS,^[9,10] DETECTIVE,^[11,12] HeMiBio,^[13–14] NOTOX,^[15–17]

[a] P. Kohonen, R. Ceder, R. C. Grafström, E. Willighagen
Karolinska Institutet
Stockholm, Sweden

[b] E. Benfenati, S. Maggioni
Istituto di Ricerche Farmacologiche Mario Negri
Milan, Italy

[c] D. Bower, M. Crump, K. Cross, S. Miller, G. Myatt
Leadscope
Columbus, USA


[d] L. Healy, G. Stacey
National Institute for Biological Standards and Control
Potters Bar, UK


[e] C. Helma, M. Rautenberg
In silico toxicology
Basel, Switzerland

[f] N. Jeliaskova, V. Jeliaskov
Ideaconsult
Sofia, Bulgaria

[g] J. Wiseman
Pharmatropé
Wayne, USA

[h] B. Hardy
Douglas Connect
Zeiningen, Switzerland
*e-mail: Barry.Hardy@douglasconnect.com

 Supporting Information for this article is available on the WWW under <http://dx.doi.org/10.1002/minf.201200114>.

 Re-use of this article is permitted in accordance with the Terms and Conditions set out at [http://onlinelibrary.wiley.com/journal/10.1002/\(ISSN\)1868-1751/homepage/2022_onlineopen.html](http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1868-1751/homepage/2022_onlineopen.html).

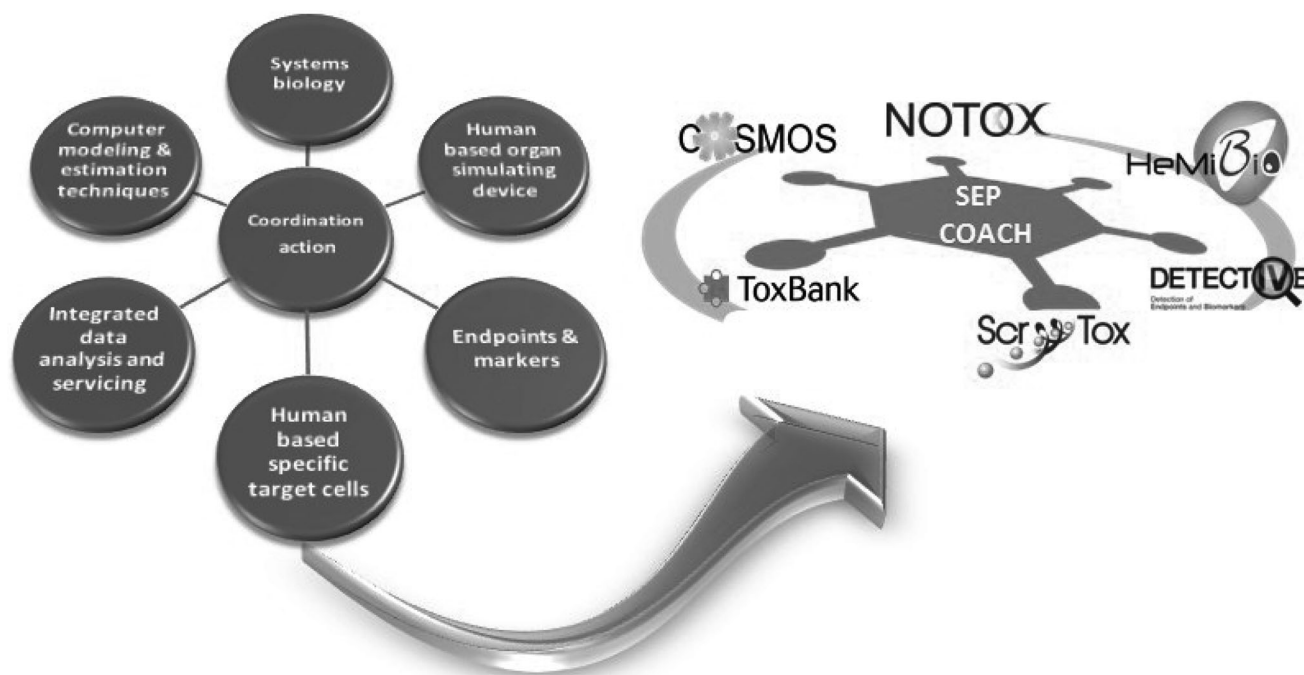


Figure 1. SEURAT-1 cluster program components and associated consortia. Design of the cluster (right) is based on the components of the EU FP7-Alternative Testing Strategies call (left).

Scr&Tox,^[18–22] an infrastructure project (ToxBank^[16,17,23]) and a coordination action project (COACH), and involves over 70 different partners (see Figure 1).

SEURAT-1 is a complex, geographically distributed, multi-disciplinary initiative. Within the SEURAT-1 cluster (HeMiBio and Scr&Tox) there is a heavy emphasis on the development of stem cell differentiation protocols to generate specific functional cell lineages as well as the use of primary cells. In addition, cell engineering methodologies are being developed to enable the use of these cells within a battery of in vitro assays, utilizing high content and high throughput automation wherever possible. An essential component of this work is establishing best practices in the use, sourcing and handling of the cells (ToxBank, Scr&Tox). Alongside this cell biology research is the development of new materials to support the use of these cells in bioreactors that mimic the 3D in vivo tissue structure (HeMiBio, NOTOX). The cluster will make extensive use of omics approaches including transcriptomics, proteomics, metabonomics, epigenetics and fluxomics coupled with systems biology approaches to understand in more detail the dynamic biological processes and to identify biomarkers that are predictive of repeated dose toxicity (DETECTIVE, NOTOX). Considerable resources are also being employed to understand biokinetics structure-activity relationships, and thresholds of toxicological concern for repeated dose toxicity and adverse outcome pathways (COSMOS). These critical technologies will be combined to create a prototype platform for safety assessment.

1.2 ToxBank

ToxBank is the cross-cluster infrastructure project whose activities include the development of the ToxBank Data Warehouse (TBDW), the selection of reference gold compounds to support the mode-of-action framework, a physical compounds repository, and resources to support the reliable use of qualified biomaterials and protocols.

ToxBank Data Warehouse. The ToxBank data warehouse provides a web-accessible shared repository of know-how and experimental results to support the SEURAT-1 cluster. The information within the TBDW is uploaded from the research activities of the cluster partners as well as relevant data and protocols from other sources, such as public databases containing toxicogenomics data.

Gold Compounds Wiki. The underlying assumption of the SEURAT-1 strategy is that we can identify MOAs that are demonstrably relevant to human toxicity based on existing knowledge such as from adverse events of marketed drugs in humans. To tackle the enormous breadth of chemical space in compound selection, we concentrate on a limited number of basic MOAs of toxicity. The SEURAT-1 goal then becomes to establish in vitro assays to characterize and represent these MOAs. Other issues such as prediction of exposure or absorption, distribution, metabolism, and excretion properties although critical to predicting human toxicity, are not determining factors per se for compound selection. A limitation of an MOA-based strategy is that our understanding of MOAs for even the best-known toxicants

is incomplete. The challenge and the opportunity are to select compounds that will enable us to increase our understanding of MOAs. Approved compound-related information is made publically available through the ToxBank wiki.^[24]

Physical compound repository. ToxBank is also responsible for the collection of all chemicals and physical properties for the evaluated standards. A framework for data quality control was followed in order to guarantee the reliability and uniformity of the collected data and their sources.

Biomaterials resources. The ToxBank cell and tissue bank will provide an important open source service to specifically enable SEURAT-1 to identify suitable sources of cell lines that will meet scientific criteria, ensure compliance with EU and national regulations and provide assays which can be taken up by industry without delays or blocks due to adverse constraints on commercial exploitation. This work utilizes standards recently developed as consensus amongst stem cell scientists and biobanks.^[25,26] The standards established will be used in ToxBank to develop evaluation criteria for suppliers of stem cell lines. Data from these suppliers will be used to demonstrate compliance with best practices.

OpenTox. In developing infrastructure, such as the data warehouse, the project is taking advantage of existing open standards, particularly the OpenTox project.^[16,17,27–31] OpenTox developed a standard framework for interoperable predictive toxicology support.^[27] It makes extensive use of REpresentational State Transfer (REST)-based web services^[32] for interaction with different geographically distributed services necessary to support predictive toxicology data management, algorithms, modeling, validation, and reporting. Extensions were made to the OpenTox framework to support additional activities needing services by ToxBank within SEURAT-1.

Open standards and the semantic web. ToxBank uses the Investigation/Study/Assay (ISA) infrastructure open source desktop software suite.^[33] Ontologies and a domain-specific ToxBank keyword hierarchy are used to enrich datasets by adding enough experimental metadata to make the archives comprehensible and reusable.^[16] The ISA2RDF tool developed by ToxBank builds on the ISA-Tab framework and facilitates conversion of investigation meta-data into the semantic web standard RDF format.

Integrated data analysis. The data in the TBDW is being collected to enable a cross-cluster integrated data analysis leading to the prediction of repeated dose toxicity within an MOA framework, based on a detailed understanding of the technologies, requirements and work practices developed across the cluster. Semantic web technologies are likely to be useful for integration of internal information from SEURAT-1 with external information from database resources around the world.^[34]

1.3 Outline of the Paper

We describe here the initial development of the ToxBank infrastructure to house all the experimental results generated from research activities carried out throughout the SEURAT-1 cluster, as well as relevant public data. The design of the system was based on a detailed understanding of the requirements and work practices used across the cluster. Technologies were selected to support the uploading and searching of protocols and experimental data, as well as future integrated data analysis needs. A series of standard reference chemicals that stratify different MOAs related to repeated dose toxicity have been selected. Anchoring the cluster activities with these chemicals supports combining the data from different projects for inference, analysis and model building in a more productive way than an unstructured approach. Finally, a case study illustrating the use of the selected compounds, the data warehouse, and analysis methods is presented.

2 Materials and Methods

2.1 Gathering Requirements for the ToxBank Data Warehouse

To build an effective solution for scientists across the SEURAT-1 cluster, it was essential to understand the work performed throughout the cluster. A methodology referred to as contextual inquiry/design was used in the collection of information for developing the requirements for the TBDW.^[35] Interviews were conducted with individual SEURAT-1 scientists and notes were taken that recorded the observations. In addition, each work task was documented, which included the ordered steps necessary to complete a task, how the event was initiated, and the reasons behind the steps. All the tasks recorded throughout all the interviews were then consolidated by common task. A second consolidation used the notes recorded from the interview to create an affinity diagram (a hierarchical view of all notes collected). The affinity diagrams and the sequences (including any consolidated sequences) were used as the starting point for the system design.

2.2 Strategy for Selection of Gold Compounds

ToxBank has created a quality-controlled curated cheminformatics database for gold compound reference standards that can be used in the training and validation of in vitro assays and in silico models for repeated dose toxicity. Gold Compound selection was carried out early in the project to facilitate decision making on project and assay design.

Reference compounds are selected primarily based on their relevance to MOA in human toxicity. However, additional criteria apply generically to all compounds to ensure their applicability for cell-based in vitro assays. These are listed in Table 1. Reference omics profiles from the literature

Table 1. Generic criteria for selecting reference compounds.

Defined, confirmed structure and isomeric form
Stable to storage, light, freeze thaw
Soluble in buffer at 30 times the in vitro IC_{50} for toxicity
Solubility in DMSO 100× buffer solubility
Insignificant binding to plasticware
Available commercially at > 95% purity (> 99% preferred)
Non-volatile
Gene expression, proteomics, metabolomics/fluxomics, and/or epigenomics profiles known
Bioactivated (hepatotoxins)

are important to this project since omics profiling will be used to characterize cellular responses to toxicants, and comparison to previously observed profiles is one mechanism for validating cellular assay systems. Criteria for acceptable physical properties ensure ease of handling in vitro assays.

2.3 Public Data

Toxicogenomics has suffered from a shortage of large publicly available standardized datasets.^[36] Recently this situation has been addressed by the release of two such datasets: the TGP (TG-GATEs) dataset^[37] and the DrugMatrix.^[38] Both have a uniform experimental design, assess a large number of marketed drugs alongside standard toxicological model compounds (such as acetaminophen) and provide in vitro as well as in vivo data for comparative analysis. Both datasets also include conventional toxicological data on the compounds for predictive modeling and phenotype anchoring. The TG-GATEs dataset includes information on 170 compounds, the DrugMatrix covers over 600 compounds and there are 73 drugs in common between the two datasets. Paired compounds – i.e., compounds that are closely related structurally but nevertheless have different toxicity profiles – comprise 16 pairs^[36] and can be used as controls to develop more accurate biomarker signatures for toxicity. SEURAT-1 gold compound selection also includes the use of paired compounds, e.g. DMNQ is structurally similar to doxorubicin but has a more specific MOA. Data from the public databases is being entered into the ToxBank data warehouse to be made available in the ISA-Tab format.

The Comparative Toxicogenomics Database (CTD) is an example of a derived resource that includes data analysis tools and curated information about gene-chemical and chemical-disease interactions that promotes understanding about the effects of environmental chemicals on human health.^[39] This strategy allows data to be integrated to construct chemical-gene-disease networks. Similar approaches can be used to summarize data from the SEURAT-1 projects and facilitate data mining and integrated data analysis.

2.4 Data Analysis

The CTD^[39] (Supporting Information SI-Table 4) was used to perform analysis of the biological similarity of gold com-

pounds, as measured by chemical-gene and chemical-gene ontology (GO) associations which were common to at least two gold compounds (Supporting Information SI 4.1–4.5). Connections to 5623 genes were assessed in the clustering analysis and 2290 GO categories were included. Clustering of the compounds by gene association was compared to clustering by GO associations and to the co-clustering of chemicals by mode-of-action (MOA) (Figure 4). The same chemical-gene and chemical – gene ontology associations were evaluated for statistically significant associations to MOA (Figure 5, Supporting Information SI Tables 5a/5b, Supporting Information SI 4.1–4.3) using the Chi-squared test.

Enrichment of gene ontology (GO) categories of genes associated with the oxidizing agent mechanisms of action (Figure 5A, Supporting Information SI Table 7) was carried out using the Webgestalt tool.^[40] Webgestalt uses the hypergeometric test to determine whether the frequency of occurrence of genes belonging to a GO-category is significantly greater in the test set than in the reference list, using all the proteins in the database with GO-annotations as background. A network of proteins associated with Asah1 protein (Figure 5C) was generated and analysis of enriched GO categories in the network was performed with the String 9.0 database (Supporting Information SI Table 6).^[41]

3 Results and Discussion

3.1 Requirements Analysis

A number of critical issues and ideas were drawn out from an analysis for the requirements data (Tables 2a and 2b). The need to develop a solution to manage the diverse protocols being generated throughout the cluster was seen as

Table 2a. Considerations in the development of the ToxBank protocols.

Protocols should cover each individual step, both the original results and any subsequent processed data
Protocols be assigned with a unique registration number for reference purposes
It should be possible to generate new versions of protocols
Protocols should cover any processing of the data (including computational steps)
Protocols may be at different stages of development (research protocol versus Standard Operating Procedure or SOP)
Sharing of protocols should be sensitive to restrictive access for intellectual property reasons or because a publication on the procedure is underway
The ability to comment on and get alerts when new protocols are uploaded
Peer-review process prior to uploading to the warehouse
Minimum effort for the end-user using any existing procedures that are currently in place
Guidelines for writing of protocols should be provided

Table 2b. Considerations in annotation of data in the ToxBank data warehouse (TBDW).

Data should always be annotated with protocols, both for the original results and any subsequent processed data
Data files should be assigned with a unique registration number
Data access should be sensitive to proprietary needs
Possible to search and download any protocol or investigation data
Both the original results and any subsequent processed data should be included
Peer-review process for data prior to uploading to the warehouse
Omics data should comply with the MIBBI minimum information standards
Data will be in the ISA-Tab format
Providing information on the cells, reagents, suppliers, and standard compounds was also highlighted as important activities
Guidelines for submitting data should be provided

a high priority as was the goal to enable SEURAT-1 investigators to upload protocols with minimum effort. Importantly, a distinction is made between research protocol and a more rigorously evaluated Standard Operating Procedure or SOP (Table 2a). Peer-review is a central component of the submission process for both protocols and data (Table 2a and 2b). Providing for restrictive access supporting the protection of intellectual property and the potential licensing of the protocols to other organizations (particularly for the SMEs in the cluster) also forms a basis for the design of the warehouse. In order to enable investigators to reproduce or judge the quality of the data as well as to support a future integrated analysis of the data, consistent information in accordance with standards such as MIBBI^[42] is collected in all laboratories for the same experiment.

Handling data presented a number of complex problems as a result of the diversity of the experiments being employed as well as the different workflows currently being adopted across the cluster (Table 2b). The system should be flexible enough to handle the diversity of data formats being generated from spreadsheets to image files. The warehouse provides access to the data from either the Graphical User Interface (GUI) or via Application Programming Interfaces (APIs) such as web services that could be incorporated within workflow management software such as KNIME.^[43]

3.2 The ToxBank Data Warehouse

3.2.1 Design of the ToxBank Data Warehouse

ToxBank manages and provides access to all protocols and experimental data across SEURAT-1 to support an integrated data analysis as illustrated in Figure 2. Once a new protocol is developed, documented and reviewed within the partner's organization, it can be uploaded to the TBDW through the ToxBank GUI where additional information is entered and associated with it.

This includes summaries of the protocol, identification of the owner and authors, and a specification of who should have access to the protocol.

In addition, keywords based on a cross-cluster keyword hierarchy, are assigned to support searching and linking. The ToxBank keyword hierarchy is a domain-specific set of searchable terms to describe the data and protocols in the TBDW (see Supporting Information SI 2.4. for details).

Investigational data generated by the research program is prepared and reviewed by SEURAT-1 investigators using the ISACreator^[33] tool to enter the experimental design, with the steps of the investigation each linked to SEURAT-1 protocols as well as to any raw or processed data. This ensures the data is in a defined and standardized format agreed across the cluster. Where an investigation is needed that is not covered by existing templates, a new investigation template is generated that is acceptable across the entire cluster. ISA-Tab archives of the investigational data are loaded into the TBDW in a similar manner as protocols.

The protocols and data can be accessed via a simple free text search or through a browse function that returns summaries of any information matching the query. The protocols or investigation data can then be viewed or downloaded directly along with links to related information, such as the ToxBank gold compound^[24] or biomaterials wikis (containing information on the gold compounds or the biomaterials being used across SEURAT-1). When the investigator does not have permission to view a specific protocol or investigation data, only the summary information is displayed. The investigator is then free to contact the investigator who loaded the information directly to the TBDW to request access rights. Once an agreement is in place, the investigator who uploaded the information modifies permission levels accordingly. A regularly scheduled email alerting scientists who have registered an interest in a specific type of information is sent out across SEURAT-1. Figure 2 summarizes the TBDW data and protocol management operations showing the phase I system that is currently implemented alongside a future phase II that will include an integrated data analysis.

3.2.2 Application of Standards to Dose-Response Analysis

The Microarray Quality Control (MAQC) consortium second phase (MAQC-II) was centered on the development and validation of genomics biomarkers and validating microarray-based models aimed at predicting toxicological and clinical end points^[44]. A key take-home message was that classifier protocols need to be more tightly described and executed. The TBDW, based on cluster-wide requirements analysis, facilitates this goal by ensuring that experimental layouts, objective metrics for data quality and conditions are captured together with SOPs for each step. For omics studies, reference dose-response analysis is being performed and carefully documented to suggest standards and protocols for analytical practice.

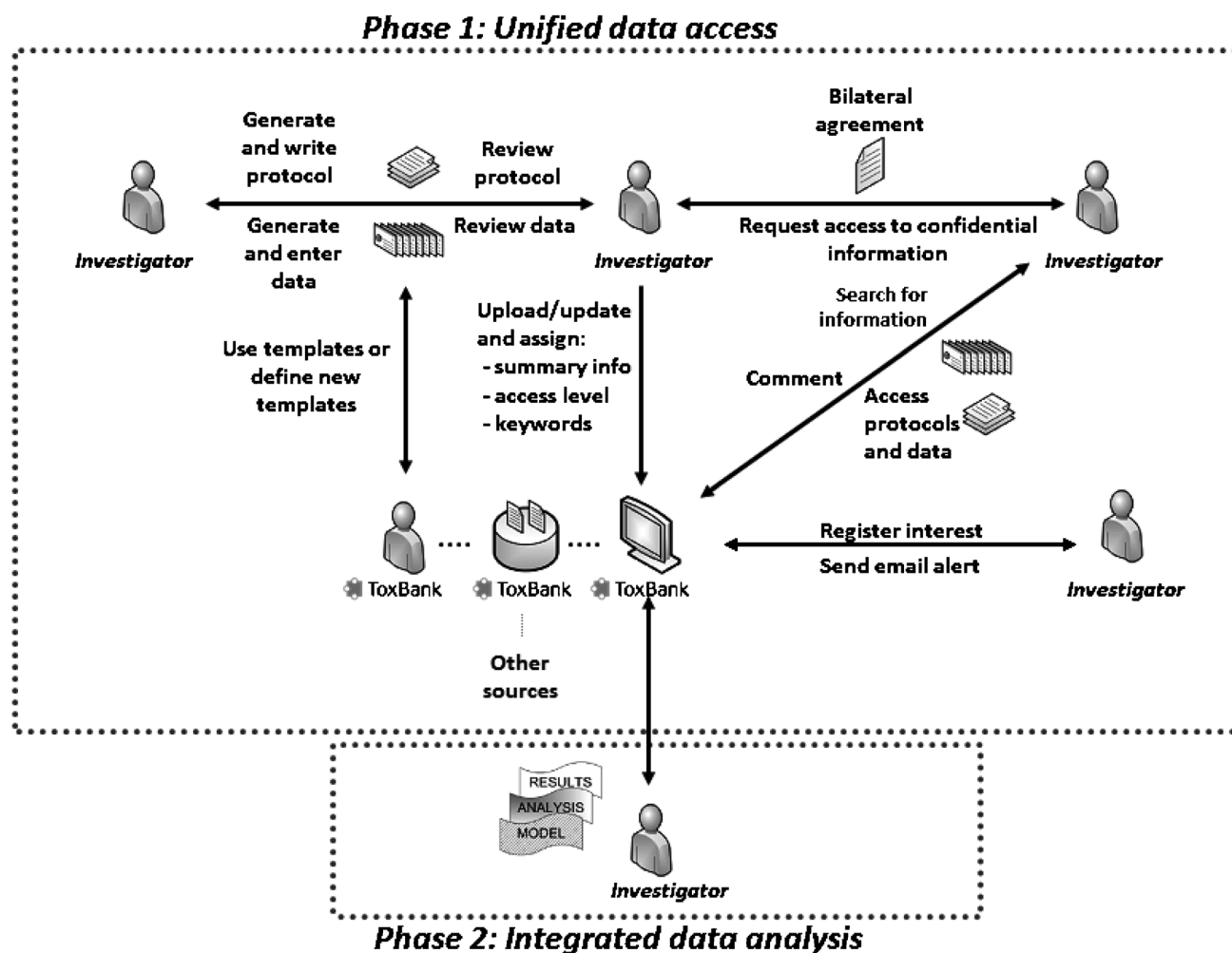


Figure 2. Overview of the currently operational (Phase I) and future (Phase II) ToxBank Data Warehouse data and protocol management operations. SEURAT-1 cluster project members share protocols, SOPs and experiment results according to bilateral agreements brokered via the TBDW. All projects upload their raw and processed data into this data warehouse using the ISA-Tab architecture.

ISA-Tab templates can store dose-response information; which together with common SOPs enables us to calculate aggregate values (e.g. IC50) from the dose-response data in a uniform way to correlate different endpoints, to identify experimental conditions having the largest impact on outcomes and to compare relative potency of compounds across different in vitro systems. Within these bounds, each consortium works towards SEURAT-1 goals independently.

3.3 ToxBank Data Warehouse Architecture and Technologies

The ToxBank system consists of a set of web services, providing access to protocols and data, a search service, and a Web GUI application, offering user-friendly access to the above functionality. The web services, developed by partners in Java and Ruby programming languages, currently share a single machine but can also be run on geographically dispersed servers, and communicate via the Internet or on a private network or secured cloud infrastructure.

This design is expected to facilitate adding new services of any kind, for example supporting different data types. ToxBank adopts the OpenTox framework design,^[27–31] based on the following technological choices (i) the REpresentational State Transfer (REST)^[32] software architecture style allowing platform and programming language independence and facilitating the implementation of new data and processing components; (ii) a formally defined common information model, based on the W3C Resource Description Framework (RDF)^[45] and communication through well-defined interfaces ensuring interoperability of the web components; (iii) authentication and authorization, allowing defining access policies of REST resources, based on OpenAM;^[46] (iv) 4store (<http://4store.org>) triple store as a backend for the investigation service. The protocol services use MySQL relational database as a backend. Both the protocol and investigation service APIs support RDF serialization of the relevant resources, regardless of the specific backend technology choice. Therefore the API and the RDF data model are inde-

pendent of the underlying implementation, enabling dynamic selection of the most appropriate backend technology, without modifying the API. The API is described at <http://api.toxbank.net/>. It is currently used by the ToxBank web GUI and an ISAcreeator extension to assess ToxBank services, and potentially could be used for ToxBank integration with other environments such as Bioclipse^[30]. The implementation of the client as well as the protocol and investigation service is open source and available at <https://github.com/ToxBank>.

ToxBank's data warehouse concept has many options for a service or business model tailored to industry members' specific requirements and needs. Choices vary from a customized deployable in-house system to a completely outsourced service solution. The distributed, modular, secure and open standards-based environment lends itself well for the needs of pharmaceuticals, agrochemicals and consumer products industrial customers, especially in the cloud-computing age when monolithic solutions are being replaced by more agile ones.

ISA-Tab is an open standard developed to provide a consistent way of representing the meta-information about an experiment.^[33] It is being used to represent the diverse types of experimental data within the TBDW. Data access and upload procedures are defined by an investigation API. Data is uploaded in ISA-Tab format and specific data queries are performed with the SPARQL query language. REST

operations are available for accessing individual investigations, studies, assays and data files and enable communication with the TBDW. To collect investigation data in the ISA-Tab format, the ToxBank consortium selected to use the ISAcreeator open source tool. The tool provides a series of forms for entering the information and can generate an archive of the entire investigation in the ISA-Tab format. The tool has been customized to integrate with resources specific to the SEURAT-1 cluster, such as users and organizations as well as SEURAT-1 protocols and common keywords.

The core standard used in the design of ToxBank to provide interoperability is, however, the Resource Description Framework.^[45] RDF is the underlying technology developed by the World Wide Web Consortium (W3C) to enable a Semantic Web.^[47] It is developed to make a Linked Resource approach possible, where information on the Internet is machine readable. RDF is the W3C proposed technology to link data on web pages, small resources or large databases, in a unified way. These Linked Resource or Semantic Web approaches are being adopted by the international life sciences community, such as the Health Care & Life Science interest group,^[48] and EU projects such as OpenTox^[27–31] and Open PHACTS.^[49] The standards support many different applications; by adopting them we ensure that ToxBank will be fully interoperable with many other life science projects. The ISA2RDF tool developed by ToxBank builds on the

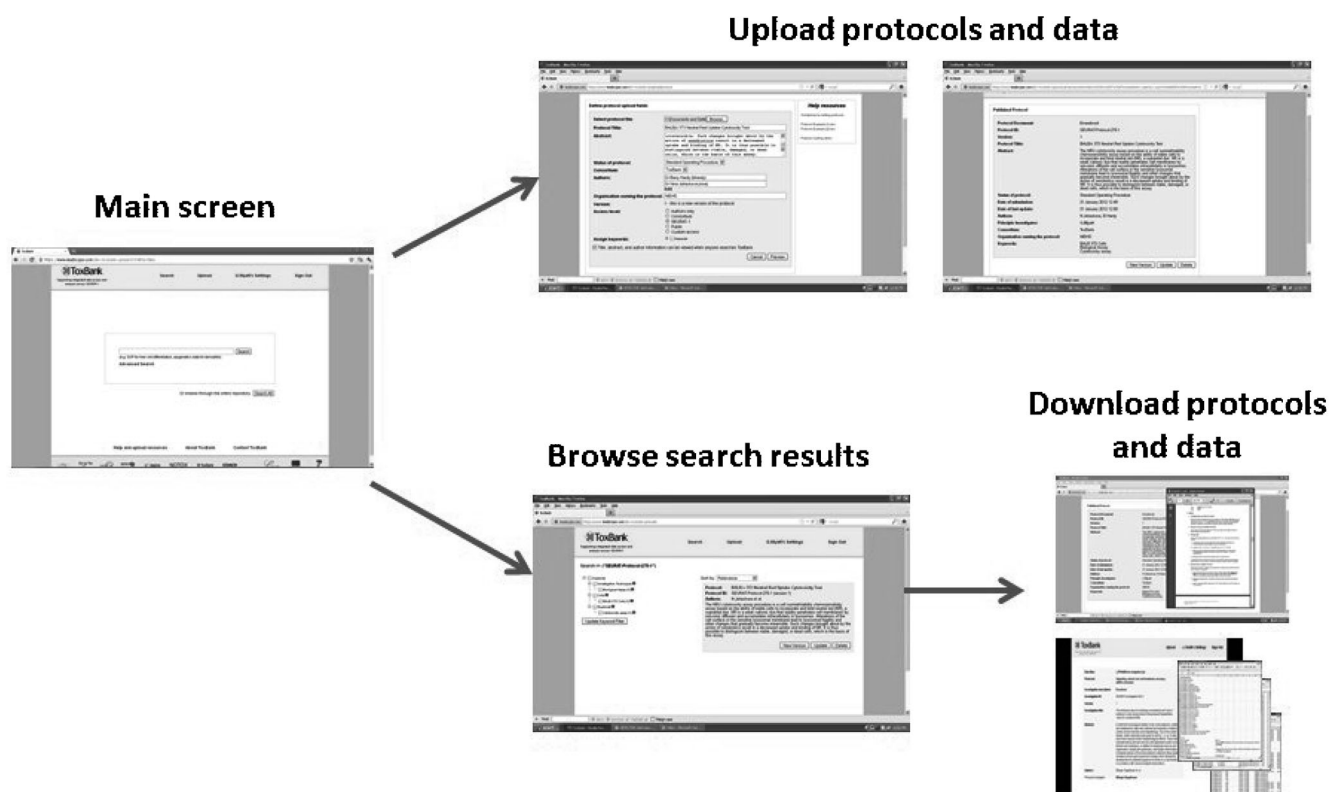


Figure 3. Screen captures of the ToxBank Data Warehouse graphical user interface for uploading, searching, browsing, and downloading cluster protocols and data (currently in the beta-testing phase), along with the ISAcreeator software^[33] for preparing an investigation dataset. Links to compound and biomaterials information in the ToxBank wiki (wiki.toxbank.net) are provided for all protocols and data sets.

ISA-Tab framework facilitating conversion of investigation meta-data into the semantic web standard RDF format (<https://github.com/ToxBank/isa2rdf>).

SEURAT-1 is a complex, multi-disciplinary initiative involving the collaboration between over 70 international partners. A common organization and definition (including any synonyms) of important concepts is an essential infrastructure enabling activity. The ToxBank consortium has created a keyword hierarchy that is used in the TBDW. In addition to its use in facilitating collaborations, the keyword hierarchy is used to support searching, browsing and linking of resources within the warehouse. When information is uploaded into the TBDW, terms will be selected from this hierarchy and linked to protocols and investigation datasets in the warehouse.

The ToxBank GUI is a front-end user interface for the repository services defined by the ToxBank API. It is a stand-alone web application allowing users to log in, review existing protocols and investigations, and to upload new protocols and investigations. The GUI serves as an intermediary between the user and the repository services. Figure 3 presents example screens from the ToxBank GUI illustrating different scenarios for uploading information and searching the content.

3.4 Gold Compound Selection

3.4.1 Overview

The proposed gold compound collection includes a limited sample of compounds for each of the major chemical mechanisms, as illustrated in Table 3. Marketed drugs with well-characterized adverse events in humans were the starting point for constructing this list. In many cases, however, the marketed drugs may have multiple activities, such as the combination of alkylating and oxidizing activities of acetaminophen. For this reason, we have added compounds that have well-characterized, narrow modes-of-action (MOA) to the collection in order to specifically represent MOAs known or presumed to commonly underlie human adverse events.

3.4.2 Alkylating Agents

Alkylating agents are distinguished in the first instance by the target nucleophile. Thiols are the cellular nucleophiles generally most reactive to alkylation, and glutathione is the most abundant thiol by several orders of magnitude. Alkylation of glutathione is frequently cited as leading to loss of cellular reduction potential, with cytotoxicity resulting from subsequent formation of free radicals.^[50] However, thiols at the active site of some proteins may be activated 100-fold or more as nucleophiles compared to glutathione.^[51] While glutathione adducts will normally be the dominant alkylated species in an absolute sense, adducts with more reactive protein thiols may be dominant in a relative sense. The balance is determined by the intrinsic reactivity of the protein thiol versus the glutathione and the efficiency of catalysis of glutathione alkylation by glutathione transferase.^[52] Existing data comparing the reactivity profiles of thiol reagents across the spectrum of cellular thiols is sparse, but reagents to quantify these profiles are available, and proteomic characterization of these profiles will be a significant contribution to the elucidation of the key molecular targets of these reagents.^[53,54]

Alkylating agents are often formed in situ by oxidation, and alkylating activity is frequently therefore associated with redox activity. This is the case for quinones such as the acetaminophen-NAPQI redox couple.^[55] In order to evaluate the effects of alkylation alone, a pure alkylating agent, iodoacetamide, has been included in the gold compound collection.

Aflatoxin B1 is included as a standard for alkylation because it is an exception in targeting lysine amino groups instead of thiols. It is activated by oxidation to an epoxide, which is subsequently hydrolyzed to a vicinal dialdehyde, which in turn forms bidentate adducts with amines.^[56,57] This toxicant also targets alkylation of DNA nitrogens, but the study of this latter reactivity is outside the current scope of the SEURAT-1 project.

3.4.3 Redox Agents

Redox agents with reduction potentials in the range between -0.3 V (NADH) and $+1.2$ V (cytochrome P450) can

Table 3. Classification of gold compounds by biochemical mechanism.

Mechanism	Representative Compounds/Receptors
Alkylating agents	
● Thiol reagents	Acetaminophen, allyl alcohol, iodoacetamide
● Lysine reagents	Aflatoxin B1
Redox agents	
● High reduction potential (strongly oxidizing)	Acetaminophen, chlorpromazine
● Low reduction potential (weakly oxidizing)	Doxorubicin
Free radical agents	Carbon tetrachloride
Promiscuous compounds	Chlorpromazine, amiodarone, valproic acid
Promiscuous receptors	Nuclear hormone receptors, hERG ion channel

be relevant to cellular biochemistry.^[58] This is an enormous range of reactivity, but the range can be split into two fundamental regions based on the reduction potential of cytochrome c. When oxidizing agents with reduction potentials lower than that for cytochrome c are reduced by NADH or glutathione, the reducing equivalents can be passed into the electron transport chain at complex III so that at least some of the energy from the reduction is preserved and converted to ATP.^[59–62] Toxicity may be assumed to arise because this process is not subject to regulation as it is in the cellular electron transport chain, and levels of NADH can be depleted below those necessary to maintain the cellular reduction potential. Direct reoxidation of the reduced form by oxygen aids in driving the cell towards an overall oxidizing environment and results in total loss of the energy stored in NADH.

In contrast, strong oxidizing agents such as NAPQI (the oxidized form of acetaminophen) trap NADH but cannot be reoxidized by cytochrome c and therefore block the entry of reducing equivalents into the electron transfer chain. This depletes the mitochondrial membrane potential and wastes the chemical energy stored in the NADH. Reoxidation of these agents requires strong oxidizing systems such as the cytochrome P450's so that toxicity is limited to metabolizing organs such as liver and kidney.^[55,63,64]

3.4.4 Free Radical Agents

Many alkylating and redox agents have been shown to induce lipid oxidation, commonly taken as evidence for free radical formation, which often arises indirectly from depletion of glutathione and NADH. The relative importance of 1-electron vs. 2-electron reactivity has not been assessed in most cases, however. We have selected carbon tetrachloride as a standard because the initial reactive species has been shown to be the trichloromethyl free radical.^[65–67]

Although the initiating factor is a free radical, a major product of this reactivity is the oxidation of fatty acids with the formation of 4-hydroxynonenal, which is a classic thiol alkylating reagent.^[67,68] The question concerning the role of free radicals in toxicity, therefore, is to what extent one- and two-electron reactivities affect different cellular targets. It must be considered, however, that differences in the effects of carbon tetrachloride compared to simple alkylating agents such as iodoacetamide may be more related to the extreme hydrophobicity of carbon tetrachloride and the reactive species derived from it.

3.4.5 Promiscuous Compounds

Chlorpromazine and amiodarone bind to phospholipid bilayers, which is the source of the pharmacological activity for these drugs.^[69,70] They are also relatively potent inhibitors of ATP synthase, which is a membrane-bound protein.^[71] While reactive metabolites are produced by chlorpromazine, amiodarone is not an obvious source of chemi-

cal reactivity. Tamoxifen, developed as an estrogen receptor (ER) antagonist, is also an ATP synthase inhibitor without obvious chemical reactivity; and both tamoxifen and amiodarone inhibit additional points in the electron transport chain, even though the proteins inhibited are disparate in chemical structure.^[71] These results together imply a role for membrane disruption by highly hydrophobic compounds in cytotoxicity, and oxidative phosphorylation appears to be sensitive to this type of inhibition.

We define compounds in this class as promiscuous in the sense that disruption of membrane function presumably affects multiple targets. There is some selectivity in the inhibition profiles, however. Additionally, while valproic acid is highly promiscuous in its activities at the high exposures commonly encountered for this drug, it is also a fatty acid analogue and there may be some selectivity for its inhibition of fatty acid oxidation compared to other toxicants in this class.^[72] Additionally, valproic acid is a selective histone deacetylase (HDAC) inhibitor and under consideration as a drug for cancer treatment.^[73]

3.4.6 Promiscuous Receptors

In addition to promiscuous ligands, there are promiscuous receptors that are relatively nonselective in binding ligands. The hERG ion channel is an archetypical example and is included as a target for cardiotoxicity standards.^[74] Promiscuity of other major protein classes is being assessed quantitatively in the Toxcast program of the EPA and is the basis of selecting the proteins below.^[75]

Nuclear factors Nrf2 and Hif-1 α are indicators of reductive and oxidative equivalents, respectively, available to the cell. They represent promiscuous responses in the sense that changes in the availability of redox equivalents are induced via a wide array of mechanisms, including alkylating and redox active toxicants described above.^[76–81]

Screening assays for nuclear hormone receptor activation also demonstrate the high promiscuity of several of these systems. Based on the profiles below, we propose to target the CAR, PXR, LXR, and AhR nuclear hormone receptors for characterization based on a common theme of regulation of lipid and steroid metabolism in hepatocytes.^[75] Additionally, to responding to xenobiotics, these receptors have major roles in cholesterol, bile acid, and fatty acid homeostasis. Compounds under consideration are listed in Table 4.

3.4.7 Energy Metabolism

Disruption of cellular energy status via loss of reduction potential, disruption of mitochondrial membrane gradients, and inhibition of ATP formation are commonly cited as causes of cytotoxicity.^[50,55] Toxicities associated with drugs, however, frequently have multiple potential MOAs, whereas, as pointed out above, an MOA-based approach to prediction of toxicity must rely on an understanding of discrete MOAs. Therefore we have identified additional com-

Table 4. Nuclear hormone receptor (NHR)-directed standards (tentative^[a]).

Compound	NHR	Comment
β -Naphthoflavone	AHR	DRE-dependent induction of metabolizing enzymes and CAR with repression of cholesterol biosynthesis.
3',4'-Dimethoxy- α -naphthoflavone	AHR	DRE-independent repression of cytokine mediated acute phase response.
GNF361	AHR	Antagonist of both DRE-dependent and -independent activity.
Phenobarbital	CAR	Induction of metabolizing enzymes and epigenetic alterations of DNA. Mechanism of CAR activation not clear.
Rifampicin	PXR	Induction of metabolizing enzymes. Gene expression profiles for human hepatocytes available.
T0901317	LXR	LXR α/β nonselective agonist induces lipogenesis, steatosis, and secretion of LDL.

[a] The compound selection strategy for NHRs is still under discussion.

pounds that are selective for key points of interaction in the energy metabolism pathways so that we can create profiles representative of these more discrete MOAs. Comparison of these to less selective toxicants will help us understand which MOAs are dominant causes of toxicity for the more complex compounds.

The thiol of glyceraldehyde phosphate dehydrogenase is highly reactive to alkylating agents, and inactivation of this enzyme would be fundamental to production of ATP and cellular survival with glucose as energy source.^[82] Thus, we have included iodoacetamide as a reference compound that is a simple thiol reagent without additional redox activity and has been well-characterized as an inhibitor of glycolysis.^[51,83]

Similarly, DMNQ is proposed as a standard for redox cycling for comparison to doxorubicin without the additional DNA intercalating activity of the latter compound, which complicates interpretation of MOA of toxicity. As described above, quinones such as DMNQ and doxorubicin cause the simple two-electron oxidation of NADH, without accompanying alkylating activity.^[61] This reaction in turn depletes the cellular reduction potential, turning on redox-sensing receptors such as Nrf2.^[80] At high concentrations, it is possible that these oxidizing agents will cause the electron transport chain to run in reverse and deplete ATP directly, as is observed for mitochondrial uncouplers.^[84]

In contrast to quinones with low reduction potentials, strongly oxidizing quinones such as the acetaminophen-NAPQI couple trap NADH but cannot be oxidized by cytochrome c and therefore block the entry of reducing equivalents into the electron transfer chain. This depletes the mitochondrial membrane potential, the cellular redox potential, and the energy stored in NADH.^[63] A central question for understanding MOAs of cytotoxicity is whether strong oxidizing agents are intrinsically different from weak oxidizing agents at the point at which NADH levels are depleted.

Finally ATP synthase is a common target of hydrophobic toxicants with promiscuous activity, presumably reflecting a sensitivity of this enzyme to membrane disruption.^[71]

3.4.8 Lipid Metabolism

Compounds with well-characterized effects on lipid metabolism in humans are commonly associated with toxicities of chemically reactive or promiscuous nature, which obscures the evaluation of phospholipidosis and cholestasis, for example, as protective, as an additional toxicity, or as a benign reaction to a xenobiotic. Thus, we have selected additional compounds that have minimum complicating associated relativities, with the purpose of assessing the relevance of long-term exposure to accumulated lipids in human toxicity. To the extent that lipid accumulation turns out to be a benign adverse event, these standards will be negative controls.

These chemically non-reactive standards include fluoxetine, a serotonin reuptake inhibitor that causes phospholipidosis by physical association with phospholipids.^[85] Bosentan is an endothelin receptor antagonist that was selected for causing cholestasis via competitive inhibition of the bile salt export pump (BSEP).^[86] Finally, dirlotapide is a compound designed to block uptake of fatty acids in the gut by inhibition of the Microsomal Triglyceride Transfer Protein (MTTP). Inhibition of hepatocyte MTTP by dirlotapide can be used to induce steatosis.^[87]

3.4.9 Tissue Repair and Fibrosis

Acetaminophen is generally considered a very safe drug on repeated low dose exposure but can progress rapidly to liver failure at exposures above a safe threshold. This is consistent with classic models of tissue repair in toxicant-induced tissue injury in which injury progresses to organ failure when the capacity for repair is exceeded.^[88] A fundamental question for prediction of repeated dose toxicity is why there is no fibrotic response to necrosis for this compound at doses below this safe threshold whereas other chemically reactive cytotoxins such as CCl₄ and allyl alcohol do cause fibrosis.^[89,90]

As an aid to understanding the causes of fibrosis independent of the complicating effects of chemical reactivity, methotrexate was selected as a reference pro-fibrotic compound with a well-defined MOA not related to chemical reactivity.^[91,92] Methotrexate is a dihydrofolate reductase in-

hibitor that acts primarily to block DNA synthesis by inhibiting conversion of dUMP to dTMP.^[93]

3.4.10 The Gold Compounds Wiki

The Gold Compounds Wiki (GCW) consists of reviewed information on the set of compounds that form the basis for the SEURAT-1 MOA strategy and is made publicly available for perusal and re-use through the ToxBank wiki.^[24] Each reference compound is annotated with a broad series of characteristics, driving the planning of the experiments, its use, and the successive interpretation of the results. The collection of data, structures and properties, represents a wide set of compounds, and investigators can take advantage of such a scheme to increase the number of reference compounds later on when the needs of further chemicals may appear. The TBDW entries for protocols and data entries are linked to the GCW as well as to biomaterials information, which will be made publicly available later in a similar manner.

3.5 Data Analysis of Public Information on Gold Compounds

The CTD was used to perform analysis of the biological similarity of gold compounds, as measured by gene and gene ontology (GO) associations with connections to at least two of the 12 gold compounds in the CTD.^[39] Cluster-

ing of the compounds by gene association ($n=5623$) was compared to clustering by GO-associations ($n=2290$) and to the co-clustering of chemicals by mode-of-action (Figure 4A and 4B). Structures of the reference compounds used in the data analysis are provided (Supporting Information SI 3, Figure S2).

Clustering of the 12 gold compounds by gene association produces an unstructured tree with few higher-level bifurcations (Figure 4A). In contrast, GO-association grouped the chemicals into three distinct clusters (Figure 4B). The uppermost cluster contains chemicals with diverse MOAs related to oxidation, such as beta-oxidation and compounds such as acetaminophen and sodium valproate. The middle cluster contains chemicals mostly having the thiol-reagent MOA, containing 75% of chemicals with that MOA. The lower-cluster contains both chemicals with the phospholipid binding MOA. Therefore GO associations seem to cluster compounds according to the MOA, at least in this set of chemicals. The better performance of GO-associations may be because individual GO categories are associated to more chemicals ($n=5.5$) on the average than are genes ($n=2.5$).

In addition, the specificity of association of individual genes and GO categories to literature-based MOAs (Table 5) were determined using the Chi-squared test (Table 6a and 6b, SI Tables 5a and 5b). The assignment of MOAs from the literature for the compounds was compared to each gene/

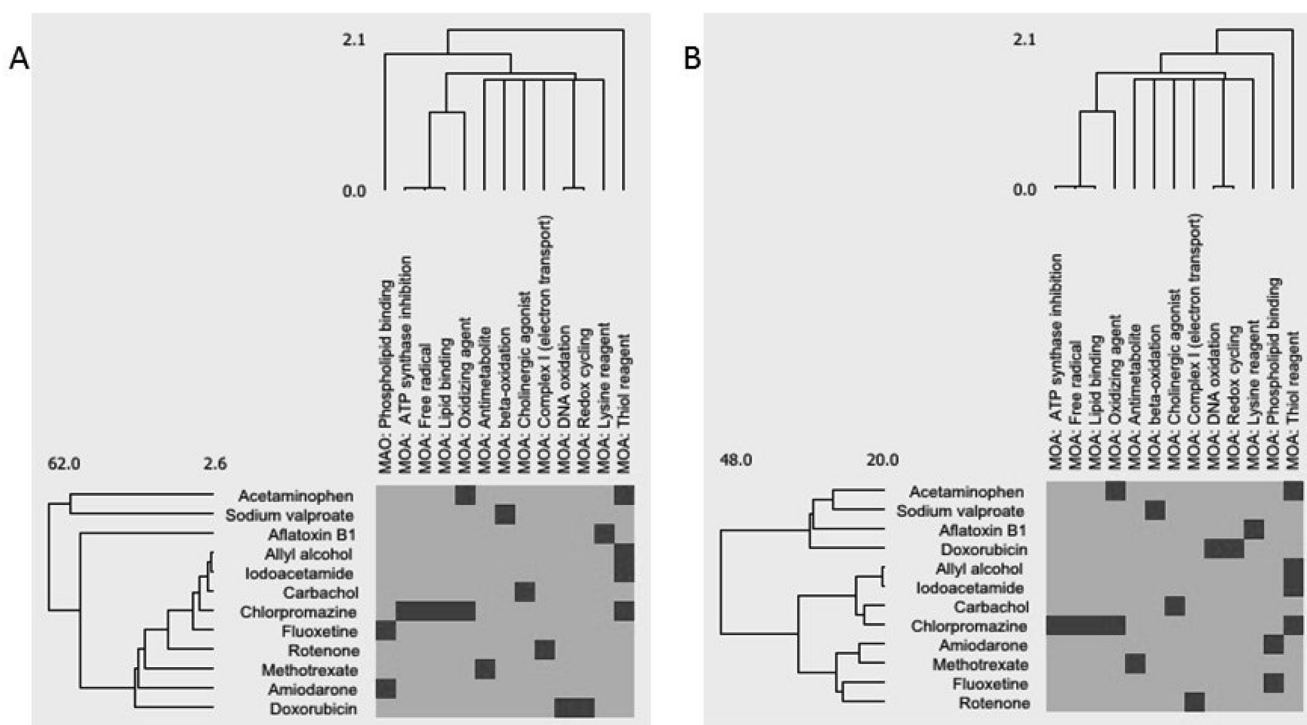


Figure 4. A) A heatmap with filled squares representing the assignment of a MOA category. Compounds are clustered according to the 5,341 genes based on links to genes in the Comparative Toxicogenomics Database (CTD) (<http://ctdbase.org/>).^[39] B) A heatmap with filled squares representing the assignment of a MOA category. Compounds are clustered according to 2290 GO categories derived based on links to genes in the CTD. The MOA categories are clustered according to the MOA profile across the compounds (Table 5).

Table 5. Summary information for reference standards. The MOAs and human adverse events for compound standards, extracted from the GCW of SEURAT-1 reference compounds. Information of the target organ(s) is also included. Compound suppliers and product numbers are provided to ensure that all labs are using a common compound source.

Compound	Target organ	MOA	Adverse event	Source and product no.
<i>Drug Standards for Reactive Compounds</i>				
Acetaminophen CAS # 103-90-2	Liver	Thiol reagent, oxidizing agent	Necrosis	Sigma Aldrich # A7085
Doxorubicin CAS # 23214-92-8	Heart	Redox cycling, DNA oxidation	Cellular lesions leading to heart failure	Sigma Aldrich # 44583
Allyl alcohol CAS # 107-18-6	Liver	Thiol reagent	Fibrosis	Sigma Aldrich # 240532
Carbon tetrachloride CAS # 56-23-5	Liver	Free radical	Fibrosis, steatosis	Sigma Aldrich # 02671
Aflatoxin B1 CAS # 1162-65-8	Liver	Lysine reagent	Apoptosis	Sigma Aldrich # A6636
Chlorpromazine CAS # 50-53-3	Liver	Thiol reagent, oxidizing agent, free radical, lipid binding, ATP synthase inhibition	Cholestasis, hepatitis	Sigma Aldrich # 31679
Iodoacetamide CAS # 144-48-9	All	Thiol reagent	(MOA standard)	Sigma Aldrich # I1149
DMNQ CAS # 6956-96-3	All	Redox cycling	(MOA standard)	Sigma Aldrich # D5439
<i>Promiscuous Ligands and Receptors</i>				
Sodium valproate CAS # 99-66-1	Liver	Inhibition of multiple pathways, including β -oxidation	Steatosis, necrosis	Sigma Aldrich # P4543
Amiodarone CAS # 1951-25-3	Liver	Phospholipid binding	Steatosis, necrosis, phospholipidosis	Tocris Bioscience # 4095
E 4031 CAS # 113558-89-7	Heart	hERG channel blocker	Arrhythmias	Sigma Aldrich # M5060
<i>MOA Standards for Oxidative Phosphorylation</i>				
Rotenone CAS # 83-79-4	All	Complex I (electron transport)	(MOA standard)	Sigma Aldrich # 45656
Oligomycin CAS # 1404-19-9	All	ATP synthase inhibitor	(MOA standard)	Tocris Bioscience # 4110
FCCP CAS # 370-86-5	All	Proton gradient uncoupler	(MOA standard)	Tocris Bioscience # 0453
<i>MOA Standards for Lipid Metabolism</i>				
Bosentan CAS # 147536-97-8	Liver	BESP inhibition	Cholestasis	Sequoia Research Products # SRP02325b
Dirlotapide CAS # 481658-94-0	Liver	MTTP inhibition	Steatosis	
Fluoxetine CAS # 54910-89-3	Liver	Phospholipid binding	Phospholipidosis	Sigma Aldrich # 34012
<i>Non-MOA Based Selections</i>				
Methotrexate CAS # 59-05-2	All	Antimetabolite	Hepatic fibrosis	Sigma Aldrich # M8407
Carbachol CAS # 51-83-2	Heart	Cholinergic agonist	(used for cell line characterization)	Sigma Aldrich # C4382
(-)Isoproterenol CAS # 7683-59-2	Heart	Adrenergic agonist	(used for cell line characterization)	Sigma Aldrich # I6504
Nifedipine CAS # 21829-25-4	Heart	L-type Ca channel blocker	(used for cell line characterization)	Sigma Aldrich # N7634
Hygromycin B CAS # 31282-04-9	All	Protein synthesis inhibitor	(standard for electron microscopy)	Invivogen # ant-hg-10p

GO category assignment for the same compounds. A Chi-squared test with Yates adjustment was used to assess the association between the assigned MOAs and the genes/GO categories. The Yates adjustment was used because of the

small number of compounds in this study. Gene/GO categories associations with the assigned MOAs were rejected with p-values greater than 0.075; however, those associations with a p-value less than 0.02 were highlighted. In ad-

Table 6a. The most strongly associated genes when comparing the MOA and gene profile across the 12 compounds in the CTD database using the Chi-square test. A Yates adjusted *p*-value of <0.075 was considered tentatively significant and *p*-value <0.02 is displayed in bold-type (Supporting Information SI Table 5a).

Mode-of-Action (MOA)	Genes
Phospholipid binding	ASAH1 , PDYN
Oxidizing agent	BAG4 , CYP2C , CYP3A11 , NDUFS3 , CAR2 , CHTF18 , STS , NFATC1 , MAPKAPK2 , SLC19A1 , CCNF , DAO , COX8A , KCNN2 , FCGR2B , FOSB , ORM1 , ORM2 , MCM5 , LEP , PHYH

Table 6b. The most strongly associated GO categories when comparing the MOA and gene profile across the 12 compounds in the CTD database using the Chi-squared test (SI Table 5b). A Yates adjusted *p*-value of <0.075 was considered tentatively significant and *p*-value <0.02 is displayed in bold-type (Supporting Information SI Table 5b).

Mode-of-Action (MOA)	Gene Ontology (GO) category
Phospholipid binding	GO:0006813 (potassium ion transport), GO:0033695 (oxidoreductase activity, acting on CH or CH2 groups, quinone or similar compound as acceptor), GO:0034875 (caffeine oxidase activity)
Oxidizing agent	GO:0019748 (secondary metabolic process), GO:0030307 (positive regulation of cell growth), GO:0000080 (G1 phase of mitotic cell cycle), GO:0051318 (G1 phase)

dition, only MOAs that were assigned to at least two gold compounds were used in this analysis. This resulted in signatures for two MOAs: phospholipid binding and oxidizing agent.

Connections of the genes signatures associated with a particular MOA (Table 6a) to gene ontology categories were identified (Figure 5A). Gene ontology (GO) category enrichment analysis of the genes associated with the oxidizing agent MOA (Table 6a) using the Webgestalt tool^[40] revealed an overabundance of genes associated with the GO molecular function (MF) Oxidoreductase activity (GO:0055114, adjP < 0.0175) and Electron carrier activity (GO:0009055, adjP < 0.0175).

Two genes are associated with the phospholipid binding MOA (Table 6a, Supporting Information SI Table 5a) and the *asah1* (*N*-acylsphingosine amidohydrolase, Yates adjusted Chi-squared *p*-value < 0.02) gene has a lipid-related function. The STRING 9.0 protein interaction and associations database^[41] was used to construct a network of directly associated proteins around Asah1 protein. A GO enrichment analysis of the proteins in the Asah1 network reveals en-

richment of the GO sphingoid metabolic process (GO:0046519, FDR *p*-value < 2.67×10^{-14}) (Figure 5C).

The GO categories associated with MOAs seem to be less mechanistically related to the MOA that was used for deriving them, although many appear to be biologically relevant (Table 6b, SI Table 4b). However, the Chi-squared *p*-values of GO-associations are also lower than the associations with genes, possibly because each GO category is, on the average, associated with a larger number of chemicals in the set of gold compounds. A difference in the average number of associations of 5.5 to 2.5 can be significant because, in the analysis, the absence of associations to non-MOA related chemicals is equally significant to the associations to chemicals with the target MOA. A more comprehensive study with a larger set of compounds would help to further define thresholds of statistical significance and best practices for connecting MOAs to genes and pathways.

Bar plots of the frequency of association of the *dao* (D-amino-acid oxidase) gene encoding a peroxisomal protein with chemicals reveals SEURAT-1 gold compounds among the top associated compounds (Figure 5B). Pretreatment of mice with the peroxisome proliferator clofibrate (CFB) protects against Acetaminophen (APAP)-induced hepatotoxicity.^[94] The protective mechanism is thought to occur via the activation of the nuclear peroxisome proliferator activated receptor- α (PPAR α). PPAR α affects the reaction of clofibrate co-treated with Acetaminophen which affects the expression of *dao* mRNA. The *dao* gene is also annotated with the enriched oxidoreductase activity GO molecular function (Figure 5A), corresponding with Acetaminophen MOA related to oxidizing agent activity.

This illustrates how association of the oxidizing agent MOA-related genes with the oxidoreductase activity GO category confirms that at least this MOA can be retrieved by analyzing the high-throughput data submitted to the CTD for the 12 compounds. The data consists mainly of unbiased genome-wide gene expression studies.

The *asah1* gene, most significantly associated with the phospholipid binding MOA, is related to the synthesis and degradation of ceramide into sphingosine and fatty acid, also showing recovery of the association to the MOA. The two gold compounds associated with *asah1* are fluoxetine and aminoadron (Figure 5D). Of these chemicals aminoadron also causes steatosis and necrosis of the liver. Differences in the associations to these chemicals may help explain why aminoadron is more toxic. One of the goals of SEURAT-1 integrated data analysis will be to find out what biomarkers signal the adverse outcomes and to understand why chemicals with a similar MOA sometimes behave differently in vivo. It is possible that in vivo repeated-dose adverse outcome prediction needs to take into account the basic MOAs of the compounds and possibly also interactions among different MOAs.

The analysis results thus confirm the assumed MOAs of the compounds with phospholipid binding and oxidizing

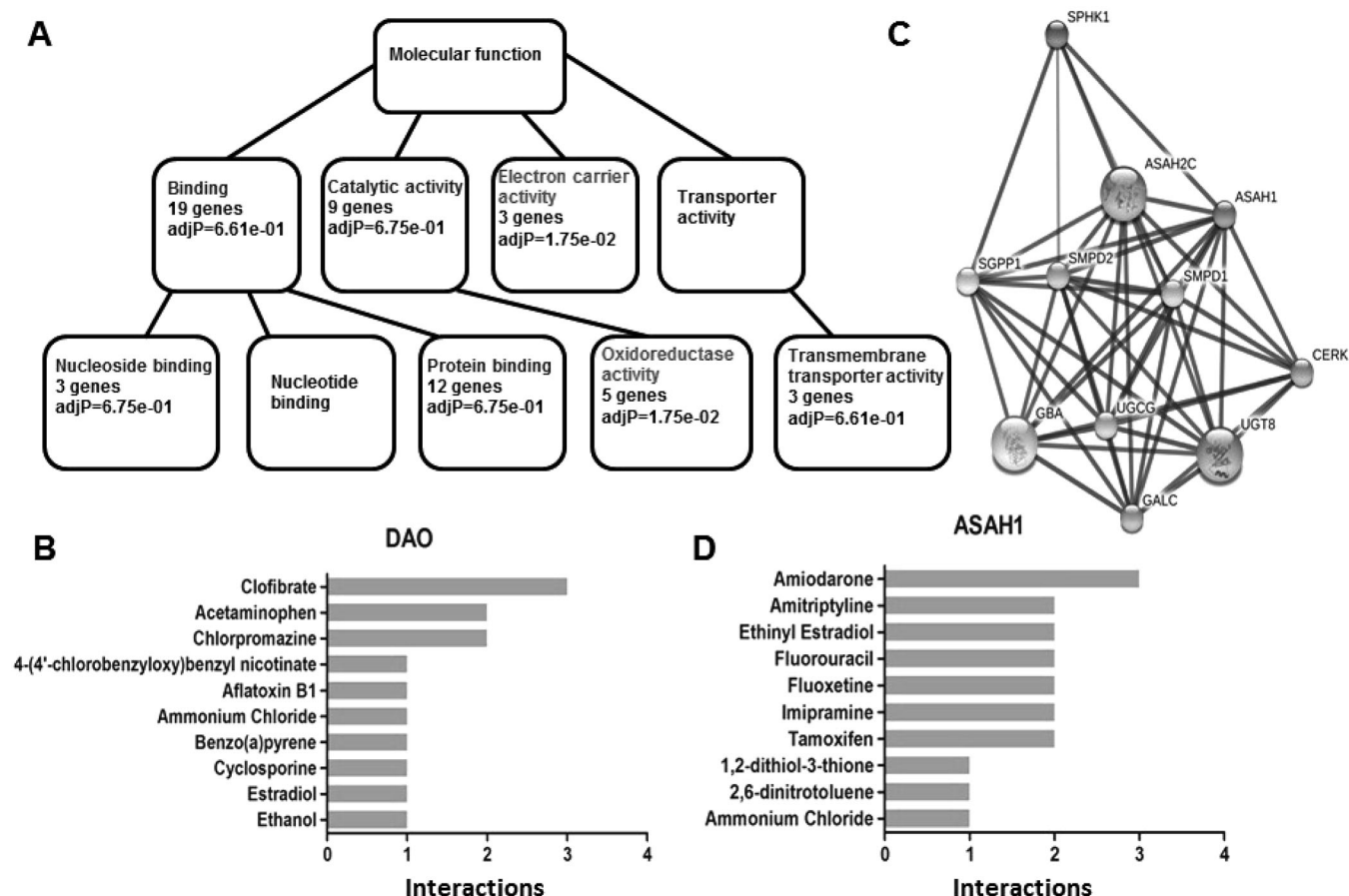


Figure 5. A) Enriched gene ontology (GO) categories of genes associated with the oxidizing agent MOA (Table 6a, SI Table 5a) using the Webgestalt tool.^[37] Terms in grey have a significance, after multiple testing correction, below $\text{adjP} < 0.05$ (Supporting Information SI Table 6). B) The gene *dao* (D-amino-acid oxidase) is associated with a number of the gold compounds according to the Comparative Toxicogenomics Database (CTD), sorted by frequency of chemical-gene associations.^[36] C) Protein-protein association network around the Asah1 (*N*-acylsphingosine amidohydrolase) protein from String 9.0. database (Supporting Information SI Table 7).^[38] D) Compounds associated with the *asah1* gene in the CTD.^[39]

agent MOAs, although these are fairly well established in the literature (Table 5). In this way the CTD derives and stores curated and statistically significant connections between chemicals, genes and GO categories; providing gene-level descriptions of the chemicals' MOAs.

4 Conclusions

The currently proposed "21st Century toxicity testing paradigm" is based on a mode-of-action framework that relies on the understanding of biological pathways and mechanisms of action that underlie the toxicity of chemicals in vivo.^[7] The SEURAT-1 program is developing a MOA-based strategy for animal-free replacements of repeated-dose toxicity testing. This paper has presented on-going work from the ToxBank infrastructure project which is supporting the research activities of the SEURAT-1 cluster, including the selection of standard reference compounds that stratify different MOAs associated with repeated dose toxicity, that

are potentially relevant across multiple endpoints and organs, such as liver, kidney, heart and the brain. The standard reference chemicals will be used within the other consortia to ensure the experimental results from the different research activities can be compared.

The ToxBank data warehouse will house SEURAT-1 generated results and protocols as well as relevant data from outside the cluster. The warehouse has been developed to enable any future integrated data analysis through the use of RDF and REST-based web services. The warehouse was designed to support research scientists working on the development of replacements to the current repeated dose toxicity tests; however, as the project develops more emphasis will be placed on the use of these approaches to support stakeholders from industry and regulatory agencies for risk assessment purposes.

A goal of the SEURAT-1 project is to investigate the applicability of model systems for uncovering chemical-MOA associations and the robustness of the associations across several model systems from 2D cultures of cell line models

to primary cell cultures to highly developed bioreactors. In order to achieve this both the descriptions of experimental metadata and the most relevant results need to be standardized with the use of ontologies and the SEURAT-1 keyword hierarchy. Semantic web technologies can enable flexible and on-going data mining of the entire dataset, as new data is generated and submitted to the ToxBank data warehouse, and can facilitate creating connections to external data such as the CTD and the processed data from the DrugMatrix and TG-GATEs repositories.

Similarly, while repeated dose toxicity is a focus for SEURAT-1, in many cases the biological rationale behind repeated dose toxicity is not fully understood. In the context of the MOA for toxicity, however, there are only two possibilities: either the MOA leading to repeated dose toxicity is the same as that for acute toxicity or it is different. To illustrate, carbon tetrachloride at high doses causes acute widespread hepatic necrosis while low repeated doses lead to fibrosis, which is still a response to necrosis, just a more limited and localized necrosis. This is an example where the primary MOA is the same for both acute and repeated dose toxicity. The repeated dose toxicity of phenobarbital, in contrast, is proposed to result from changes in locus-specific DNA methylation patterns, an MOA distinct from acute biological responses.^[8]

The compound selection strategy must therefore be based on an understanding of MOAs that underlie repeated dose toxicity so that these MOAs are adequately represented in the in vitro assays. The difficulty behind this statement is again illustrated by carbon tetrachloride. While carbon tetrachloride-induced fibrosis is presumably a response to cell death, many compounds cause hepatic cell death, but not all of them cause fibrosis upon repeated low exposures. The challenge is to understand and derive from cellular model system MOAs at a level of granularity sufficient to distinguish acute versus chronic effects in vivo.

SEURAT-1 gold compounds with diverse chemical structures (Supporting Information SI Figure S2) and MOAs (Table 5, Figure 5) were clustered by chemical-gene and chemical-GO association from the CTD. Clustering of the compounds by GO association grouped together compounds with a similar MOA. Associations of genes to the SEURAT-1 gold compounds from the CTD come mainly from microarray studies, illustrating the value of gene expression studies in generating a large number of chemical-gene associations. Microarray studies also generate more unbiased chemical-gene associations, since typically the expression levels of all the protein-coding genes in the genome are measured. For validation purposes associations uncovered by the toxicogenomics analysis would need to be experimentally verified.

Since the SEURAT-1 gold compounds have fairly well established MOAs (Table 5) the most significant result of the CTD-based data analysis is that unbiased high-throughput data also reflects the MOAs obtained from literature despite the relatively small number of compounds in the anal-

ysis, the diverse studies submitted to the CTD, and despite the compounds with the same MOA having different chemical structural features. Results of the SEURAT-1 high-throughput 'omics profiling experiments obtained from treatments of the cellular model systems with the gold compounds and submitted to the TBDW can be analyzed in the same fashion.

The use of the CTD to analyze MOAs relevant to SEURAT-1 gold compounds also illustrates how statistically significant chemical-gene associations can be mined at the gene-level and connected to MOAs via GO categories. Such associations, if derived using a sufficiently large set of compounds to ensure specificity, can be tentatively considered biomarkers for the detection of toxicologically relevant MOAs. MOA-specific genes and pathways that are detected across several model systems or in the highest quality ones would then form a basis for designing in vitro reporter assays, such as those that are envisioned for further developments related to SEURAT-1 and its extensions.

The ultimate goal would be to fully recreate human in vivo conditions in culture but the more realistic goal of SEURAT-1 is to work towards animal-free repeated-dose toxicity testing of chemical entities overall. Establishing cellular models that enable determination of toxicity relevant modes-of-action in a reproducible manner represents an important step on the way.

Acknowledgements

The research leading to these results has received funding from the *European Community's Seventh Framework Programme* (FP7/2007–2013) under Grant Agreement n° [267042]. The research leading to these results has received financing from *Cosmetics Europe*. We would like to thank Dr. Scott Auerbach from the NIEHS for providing information on the DrugMatrix database and Drs. Susanna-Assunta Sansone, Philippe Rocca-Serra and Eamonn Maguire for their help with the ISA-Tab and ISAcreator software. We would also like to acknowledge our appreciation of SEURAT-1 partners who collaborated on the requirements gathering, Dr. Brigitte Landesmann, and the Gold Compound Working Group.

References

- [1] SEURAT-1, available from, <http://www.seurat-1.eu/>, 2012.
- [2] *Seventh Framework Programme (FP7)*, available from, http://cordis.europa.eu/fp7/home_en.html, 2012.
- [3] *Cosmetics Europe*, available from, <http://www.cosmeticseurope.eu>, 2012.
- [4] A. R. Boobis, J. E. Doe, B. Heinrich-Hirsch, M. E. Meek, S. Munn, M. Ruchirawat, et al., *Crit. Rev. Toxicol.* **2008**, 38, 87–96.
- [5] G. T. Ankley, R. S. Bennett, R. J. Erickson, D. J. Hoff, M. W. Hornung, R. D. Johnson, et al., *Environ Toxicol Chem.* **2010**, 29, 730–41.
- [6] K. Boekelheide, M. E. Andersen, *ALTEX* **2010**, 27, 243–52.

- [7] D. Krewski, D. Acosta Jr., M. H. Andersen, Anderson, J. C. Bailar, 3rd, K. Boekelheide, et al., *J. Toxicol. Environ. Health B Crit. Rev.* **2010**, 13, 51–138.
- [8] T. Gocht, M. Schwarz, *SEURAT-1 – Towards the replacement of in vivo repeated dose systemic toxicity testing*, Vol. 2, ISBN , 978-2-9539634-1-0, Coach consortium, **2012**.
- [9] COSMOS, available from, <http://www.cosmos-tox.eu/>, **2012**.
- [10] A. Tonnelier, S. Coecke, J. M. Zaldivar, *Arch Toxicol.* **2012**, 86, 393–403.
- [11] DETECTIVE, available from, <http://www.detect-iv-e.eu/>, **2012**.
- [12] M. Wiesinger, B. Mayer, P. Jennings, A. Lukas, *Toxicol In Vitro.* **2012**, 26, 956–62.
- [13] HeMiBio, available from, <http://www.hemibio.eu/>
- [14] M. Vinken, M. Pauwels, G. Ates, M. Vivier, T. Vanhaecke, V. Rogiers, *Arch Toxicol.* **2012**, 86, 405–12.
- [15] NOTOX, available from, <http://www.notox-sb.eu/>, **2012**.
- [16] B. Hardy, G. Apic, P. Carthew, D. Clark, D. Cook, I. Dix, et al., *ALTEX* **2012**, 29, 129–37.
- [17] B. Hardy, G. Apic, P. Carthew, D. Clark, D. Cook, I. Dix, et al., *ALTEX* **2012**, 29, 139–56.
- [18] Screentox, available from, <http://www.scrtox.eu/>, **2012**.
- [19] S. Colleon, C. Galli, J. A. Gaspar, K. Meganathan, S. Jagtap, J. Hescheler, et al., *Toxicol. Sci.* **2011**, 124, 370–377.
- [20] P. Sartipy, P. Bjorquist, *Stem Cells* **2011**, 29, 744–748.
- [21] F. Pistollato, S. Bremer-Hoffmann, L. Healy, L. Young, G. Stacey, *Expert. Opin. Drug Metab. Toxicol.* **2012**, 8, 239–57.
- [22] J. Synnergren, C. Ameen, A. Jansson, P. Sartipy, *Physiol. Genomics.* **2012**, 44, 245–58.
- [23] ToxBank, available from, <http://toxbank.net/>, **2012**.
- [24] Gold Compounds Wiki (GCW), available from, <http://wiki.toxbank.net>, **2012**.
- [25] S. Coecke, M. Balls, G. Bowe, J. Davis, G. Gstraunthaler, T. Hartung, et al., *Altern. Lab. Anim.* **2005**, 33, 261–287.
- [26] International Stem Cell Banking Initiative, *Stem Cell Rev.* **2009**, 5, 301–314.
- [27] OpenTox, available from, <http://www.opentox.org/>, **2012**.
- [28] B. Hardy, N. Douglas, C. Helma, M. Rautenberg, N. Jeliaskova, V. Jeliaskov, et al., *J. Cheminform.* **2010**, 2, 7.
- [29] N. Jeliaskova, V. Jeliaskov, *J. Cheminform.* **2011**, 3, 18.
- [30] E. L. Willighagen, N. Jeliaskova, B. Hardy, R. C. Grafstrom, O. Spjuth, *BMC Res. Notes.* **2011**, 4, 487.
- [31] O. Tcheremenskaia, R. Benigni, I. Nikolova, N. Jeliaskova, S. E. Escher, M. Batke, et al., *J. Biomed. Semantics* **2012**, 3 Suppl 1, S7.
- [32] RESTful webservices, R. T. Fielding, R. N. Taylor, *ACM Transactions on Internet Technology (TOIT)*, New York, Association for Computing Machinery, **2002–05**, 2 (2), 115–150.
- [33] S. A. Sansone, P. Rocca-Serra, D. Field, E. Maguire, C. Taylor, O. Hofmann, et al., *Nat Genet.* **2012**, 44, 121–126.
- [34] D. J. Wild, Y. Ding, A. P. Sheth, L. Harland, E. M. Gifford, M. S. Lajiness, *Drug Discov. Today* **2012**, 17, 469–474.
- [35] K. Holtzblatt, J. Burns Wendell, S. Wood, *Rapid Contextual Design, A How-to Guide to Key Techniques for User-Centered Design*, Elsevier, **2005**.
- [36] M. Chen, M. Zhang, J. Borlak, W. Tong, *Toxicol. Sci.* **2012**.
- [37] T. Uehara, A. Ono, T. Maruyama, I. Kato, H. Yamada, Y. Ohno, et al., *Mol. Nutr. Food Res.* **2010**, 54, 218–227.
- [38] B. Ganter, S. Tugendreich, C. I. Pearson, E. Ayanoglu, S. Baumhueter, K. A. Bostian, et al., *J. Biotechnol.* **2005**, 119, 219–244.
- [39] A. P. Davis, B. L. King, S. Mockus, C. G. Murphy, C. Saraceni-Richards, M. Rosenstein, et al., *Nucleic Acids Res.* **2011**, 39, D1067–72.
- [40] B. Zhang, S. Kirov, J. Snoddy, *Nucleic Acids Res.* **2005**, 33, W741–748.
- [41] D. Szklarczyk, A. Franceschini, M. Kuhn, M. Simonovic, A. Roth, P. Minguez, et al., *Nucleic Acids Res.* **2011**, 39, D561–568.
- [42] C. F. Taylor, D. Field, S. A. Sansone, J. Aerts, R. Apweiler, M. Ashburner, et al., *Nat. Biotechnol.* **2008**, 26, 889–896.
- [43] W. A. Warr, *J. Comput. Aided Mol. Des.* **2012**, 26, 801–804.
- [44] MAQC Consortium, *Nat. Biotechnol.* **2010**, 28(8), 827–838.
- [45] RDF, J. J. Carroll, G. Klyne, “Resource Description Framework (RDF), Concepts and Syntax”, *Technical Report, W3C*, **2004**.
- [46] OpenAM, available from, <http://openam.forgerock.org>, **2012**.
- [47] Semantic Web, available from, <http://semanticweb.org>, **2012**.
- [48] M. Samwald, A. Jentzsch, C. Bouton, C. S. Kallesøe, E. Willighagen, J. Hajagos, M. S. Marshall, E. Prud'hommeaux, O. Hassenzadeh, E. Pichler, S. Stephens, *J. Cheminform.* **2011**, 3(1), 19.
- [49] A. J. Williams, L. Harland, P. Groth, S. Pettifer, C. Chichester, E. L. Willighagen, et al., *Drug Discov. Today.* **2012**, 17(21–22), 1188–1189.
- [50] M. J. Gomez-Lechon, L. Tolosa, J. V. Castell, M. T. Donato, *Toxicol. In Vitro* **2010**, 24, 1879–1889.
- [51] M. M. Schmidt, R. Dringen, *Front. Neuroenergetics* **2009**, 1, 1.
- [52] B. Coles, I. Wilson, P. Wardman, J. A. Hinson, S. D. Nelson, B. Ketterer, *Arch Biochem Biophys.* **1988**, 264, 253–260.
- [53] A. T. Jacobs, L. J. Marnett, *Acc. Chem. Res.* **2010**, 43, 673–683.
- [54] S. Hoogendoorn, L. Willems, B. Florea, H. Overkleeft, *Angew. Chem. Int. Ed. Engl.* **2011**, 50, 5434–5436.
- [55] W. Tang, *Expert. Opin. Drug Metab. Toxicol.* **2007**, 3, 407–420.
- [56] T. O'Brien, G. Babcock, J. Cornelius, M. Dingeldein, G. Talaska, D. Warshawsky, et al., *Toxicol. Appl. Pharmacol.* **2000**, 164, 280–290.
- [57] F. P. Guengerich, K. O. Arneson, K. M. Williams, Z. Deng, T. M. Harris, *Chem. Res. Toxicol.* **2002**, 15, 780–792.
- [58] W. H. Koppenol, *J. Am. Chem. Soc.* **2007**, 129, 9686–9690.
- [59] M. Floreani, F. Carpenedo, *Gen. Pharmacol.* **1995**, 26, 1757–1764.
- [60] J. M. Karczewski, J. G. Peters, J. Noordhoek, *Biochem. Pharmacol.* **1999**, 57, 27–37.
- [61] R. H. Haefeli, M. Erb, A. C. Gemperli, D. Robay, I. Courdier Fruh, C. Anklin, et al. *PLoS One* **2011**, 6, e17963.
- [62] Y. Wen, W. Li, E. C. Poteet, L. Xie, C. Tan, L. J. Yan, et al., *J. Biol. Chem.* **2011**, 286, 16504–16515.
- [63] J. A. Hinson, D. W. Roberts, L. P. James, *Handb. Exp. Pharmacol.* **2010**, 369–405.
- [64] Y. Song, G. R. Buettner, *Free Radic. Biol. Med.* **2010**, 49, 919–962.
- [65] T. Castillo, D. R. Koop, S. Kamimura, G. Triadafilopoulos, H. Tsukamoto, *Hepatology* **1992**, 16, 992–996.
- [66] L. E. Fleming, in *Case Studies in Environmental Medicine* (Eds: A. M. Pope, D. P. Rall), National Academies Press, Washington, DC, **1995**, p. 249–266.
- [67] L. W. Weber, M. Boll, A. Stampfl, *Crit. Rev. Toxicol.* **2003**, 33, 105–136.
- [68] M. Boll, L. W. Weber, E. Becker, A. Stampfl, *Z. Naturforsch. C* **2001**, 56, 649–659.
- [69] P. Seeman, *Biochem. Pharmacol.* **1977**, 26, 1741–1748.
- [70] N. Anderson, J. Borlak, *FEBS Lett.* **2006**, 580, 5533–5540.
- [71] S. Nadeanacia, A. Bernal, R. Aggeler, R. Capaldi, Y. Will, *Toxicol. In Vitro* **2007**, 21, 902–911.
- [72] J. W. Kesterson, G. R. Granneman, J. M. Machinist, *Hepatology* **1984**, 4, 1143–1152.
- [73] C. J. Phiel, F. Zhang, E. Y. Huang, M. G. Guenther, M. A. Lazar, P. S. Klein, *J. Biol. Chem.* **2001**, 276, 36734–36741.
- [74] F. De Ponti, E. Poluzzi, A. Cavalli, M. Recanatini, N. Montanaro, *Drug Saf.* **2002**, 25, 263–286.

- [75] M. T. Martin, D. J. Dix, R. S. Judson, R. J. Kavlock, D. M. Reif, A. M. Richard, et al., *Chem. Res. Toxicol.* **2010**, *23*, 578–590.
- [76] R. Tirumalai, T. Rajesh Kumar, K. H. Mai, S. Biswal, *Toxicol Lett.* **2002**, *132*, 27–36.
- [77] G. Olmos, I. Conde, I. Arenas, L. Del Peso, C. Castellanos, M. O. Landazuri, et al., *Cell Signal.* **2007**, *19*, 2098–2105.
- [78] I. M. Copple, C. E. Goldring, R. E. Jenkins, A. J. Chia, L. E. Randle, J. D. Hayes, et al. *Hepatology* **2008**, *48*, 1292–1301.
- [79] Y. L. Chua, E. Dufour, E. P. Dassa, P. Rustin, H. T. Jacobs, C. T. Taylor, et al., *J. Biol. Chem.* **2010**, *285*, 31277–31284.
- [80] A. V. Pointon, T. M. Walker, K. M. Phillips, J. Luo, J. Riley, S. D. Zhang, et al., *PLoS One* **2010**, *5*, e12733.
- [81] E. M. Sparkenbaugh, Y. Saini, K. K. Greenwood, J. J. LaPres, J. P. Luyendyk, B. L. Copple, et al., *J. Pharmacol. Exp. Ther.* **2011**, *338*, 492–502.
- [82] N. J. Allen, R. Karadottir, D. Attwell, *J. Neurosci.* **2005**, *25*, 848–859.
- [83] J. R. Williamson, *J. Biol. Chem.* **1967**, *242*, 4476–4485.
- [84] A. L. Nieminen, A. K. Saylor, B. Herman, J. J. Lemasters, *Am. J. Physiol.* **1994**, *267*, C67–74.
- [85] H. Sawada, K. Takami, S. Asahi, *Toxicol. Sci.* **2005**, *83*, 282–292.
- [86] K. Fattinger, C. Funk, M. Pantze, C. Weber, J. Reichen, B. Stieger, et al., *Clin. Pharmacol. Ther.* **2001**, *69*, 223–231.
- [87] R. P. Robinson, J. A. Bartlett, P. Bertinato, A. J. Bessire, J. Cosgrove, P. M. Foley, et al., *Bioorg. Med. Chem. Lett.* **2011**, *21*, 4150–4154.
- [88] H. M. Mehendale, *Toxicol. Pathol.* **2005**, *33*, 41–51.
- [89] S. A. Jung, Y. H. Chung, N. H. Park, S. S. Lee, J. A. Kim, S. H. Yang, et al., *Scand. J. Gastroenterol.* **2000**, *35*, 969–975.
- [90] S. Hoehme, M. Brulport, A. Bauer, E. Bedawy, W. Schormann, M. Hermes, et al., *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 10371–10376.
- [91] P. Hytioglou, H. Tobias, R. Saxena, M. Abramidou, C. S. Papadimitriou, N. D. Theise, *Am. J. Clin. Pathol.* **2004**, *121*, 324–329.
- [92] K. Lindsay, A. D. Fraser, A. Layton, M. Goodfield, H. Gruss, A. Gough, *Rheumatology (Oxford)*, **2009**, *48*, 569–572.
- [93] J. M. Kremer, *Arthritis Rheum.* **2004**, *50*, 1370–1382.
- [94] J. S. Moffit, P. H. Koza-Taylor, R. D. Holland, M. S. Thibodeau, R. D. Beger, M. P. Lawton, et al., *Toxicol. Appl. Pharmacol.* **2007**, *222*, 169–79.

Received: October 4, 2012

Accepted: November 27, 2012

Published online: January 17, 2013